Introduction to Web Science 101
Assignment 1

Victor Nwala

Department of Computer Science

Old Dominion University

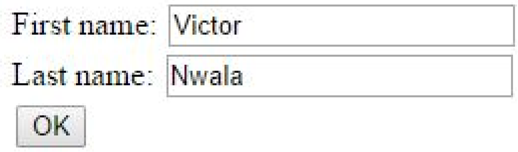
09/08/2014

CS 495/595 Assignment № 1

Victor Nwala, Old Dominion University 09/08/2014

QUESTION ONE SOLUTION

To answer this question I designed a simple html form on my localhost that looks somewhat like what we see below.



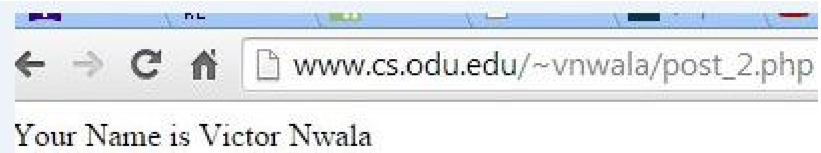
First name: Victor

Last name: Nwala

The form is submitted, to the page below.

```
<html>
<head>
</head>
<body>


<?php

 echo 'Your Name is'." ".$_POST['firstname']." ".$_POST['lastname'];

?>




</body>
</html>
```

And the output is seem below also.



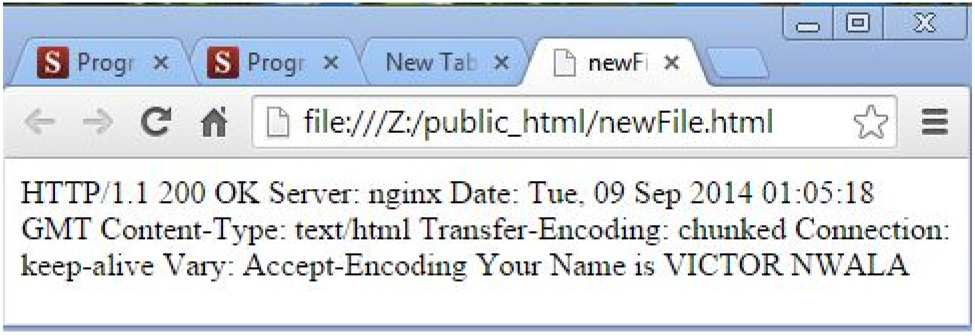
www.cs.odu.edu/~vnwala/post_2.php

Your Name is Victor Nwala

I also did the same using a curl command, I used curl -d to send parameters to my final page, I used -i to include the html response and -o to output the result into a new file (newFile.html). The command looks like this:

```
curl -d "firstname=VICTOR&lastname=NWALA&press=OK"-i
 www.cs.odu.edu/~vnwala/post_2.php -o "newFile.html"
```

I opened the saved file in my browser and the result is seem below:



HTTP/1.1 200 OK Server: nginx Date: Tue, 09 Sep 2014 01:05:18 GMT Content-Type: text/html Transfer-Encoding: chunked Connection: keep-alive Vary: Accept-Encoding Your Name is VICTOR NWALA

The image above display the output and includes the html response.

QUESTION TWO SOLUTION

Python program that receives three arguments, school name, time and url and gives the current scores.

The requirements of the program are to: a)Receive three arguments b)Run continuously until interrupted by hitting crtl c on the keyboard.

To explain what I did I will highlight some import steps in my python code below.

Line 8, 9 and 31 ensures the function runs infinitely.

Line 11, with the help of BeautifulSoup and urllib2 library, the url is opened and stored as a huge string.

Line 12, the table rows of class game link which contains the scores are located and stored. Subsequently, the table data of classes home and away are stored also.

Line 26 and 28 ensure that scores will only be printed if the school name entered is located.

Line 30 ensures that the program sleeps for the number of seconds received in the argument.

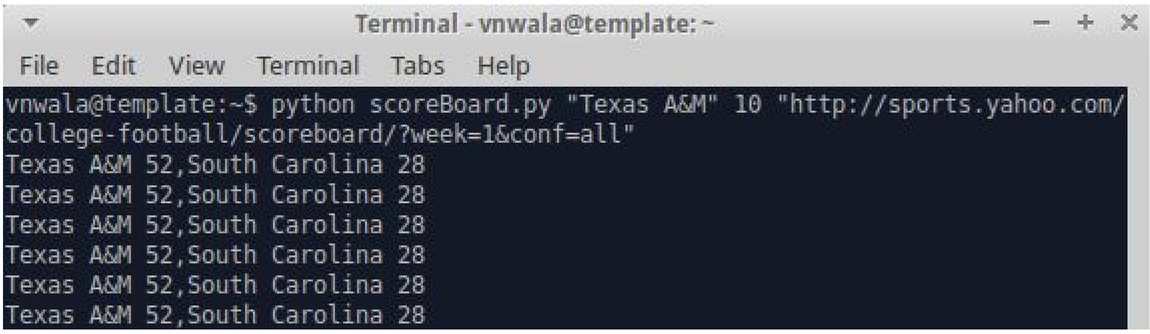
Line 33 and 34 call the parent function only if the arguments are entered properly.

```
1 import time
2 import sys
3 import urllib2
4 from bs4 import BeautifulSoup
5
```

```

def extractScores(school, seconds,url):
        count = 1
        while ( count > 0):
        seconds = float (seconds)
        soup = BeautifulSoup(urllib2.urlopen(url).read())
        for row in soup('table')[1].findAll('tr', {'class':'game  link'}):
            away = row.find('td', {'class':'away'})
            away = away.em.text

            awayScore = row.find('td', {'class':'score'}).findAll('span')
            awayScore = awayScore[0].text

            home = row.find('td', {'class':'home'})
            home =  home.em.text


            homeScore = row.find('td', {'class':'score'}).findAll('span')
            homeScore = homeScore[1].text

                if (away.lower() == school.lower() or home.lower() == school.lower()):

                print away + " "+ awayScore + "," + home+ " " + homeScore

                    time.sleep(seconds)
        count = count + 1

if len(sys.argv) == 4:
    extractScores(sys.argv[1],sys.argv[2],sys.argv[3])
else:
    print "Error, wrong argument count, do this instead: ", sys.argv[0] + " schoolName, url and time"
```

I entered the school name Texas A&M, and a given url and asked to program to sleep for 10 seconds before running again, as shown below.

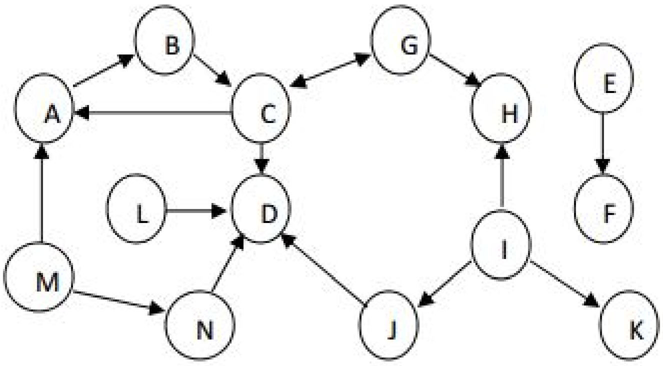
```
vnwala@template:~$ python scoreBoard.py "Texas A&M" "http://sports.yahoo.com/college-football/scoreboard/?week=1&conf=all" 10
```



```
vnwala@template:~$ python scoreBoard.py "Texas A&M" 10 "http://sports.yahoo.com/
college-football/scoreboard/?week=1&conf=all"
Texas A&M 52,South Carolina 28
Texas A&M 52,South Carolina 28
Texas A&M 52,South Carolina 28
Texas A&M 52,South Carolina 28
Texas A&M 52,South Carolina 28
Texas A&M 52,South Carolina 28
```

I should state at this point that a fellow classmate, Alexander Nwala, help me with the logic of my Python Program.

QUESTION THREE SOLUTION



DEFINITIONS:

1. IN: Pages with no in-links or in-links from IN pages and out-links to pages in IN, SCC, Tendrils, or Tubes.

2. SCC: Pages with in-links from IN or SCC and out-links to OUT or SCC. There exists some path of links from every page in SCC to every other page in SCC.

3. OUT: Pages with no out-links or out-links to other pages in OUT, and all in-links come from OUT, SCC, Tendrils, or Tubes.

4. Tendrils: Pages that can only be reached from IN or have only out-links to OUT.

5. Tubes: Pages that have in-links from IN or other pages in Tubes and out-links to pages in Tubes or OUT.

6. Disconnected: Pages that have no in-links from any other components and no out-links to other components. These pages may be linked to each other.

SOURCE:

http://www.harding.edu/fmccown/classes/comp475-s13/web-structure-homework.pdf

From the graph above, the pages (links), A, B, C, G can reach every other member of this group. Hence they consist of the SCC component of the Bow-Tie web Structure.

The page M has no in-links and out-links to pages in SCC, hence it can be called the IN component of the Bow-Tie web Structure.

The pages D and H have no out-links and its in-links come from SCC, hence its consists of the OUT component of the Bow-Tie web Structure.

The pages, L, I, J have only out-links to OUT, hence the are categorized as Tendrils, K is also categorized as a Tendril because it is connected to I.

The page N, has an in-link from IN (M) and out-links to an OUT page (D), hence N is categorized as a TUBE.

The pages E and F have no in-links from any other components and no out-links to other components, hence they are called the DISCONNECTED component of the Bow-Tie Web Structure.

SCC: A,B,C,G

IN: M

OUT: D,H

TENDRILS: L,I,J,K

TUBE: N

DISCONNECTED: E,F